CLAYTONE SIKASOTE

https://csikasote.github.io

claytonsikasote@gmail.com | +260-979-084696

RESEARCH INTEREST

I'm interested in artificial and machine intelligence with focus on language and speech processing. In particular, I'm enthusiastic about building efficient and robust speech recognition systems in low-resource and multilingual settings. I'm also interested in understanding the intricacies of human-machine interactions towards building intelligent multimodal dialogue systems. Presently, my research is focussed on exploring various approaches towards curating high-quality speech and language data resources that would enable research and development of language technology systems for Zambian languages.

EDUCATION

Phd., Computer Science, University of Cape Town, Cape Town, South Africa   Feb 2024 – **Present**
  - Supervisor(s): Dr. Jan Buys and Prof. Hussein Suleman, University of Cape Town, South Africa

MSc., Mathematical Sciences, Africa Institute of Mathematical Sciences, Kigali, Rwanda   Oct 2019 – Aug 2020
  - Supervisor: Prof. Antonios Anastasopoulos, George Mason University, U.S.A
  - Specialization: **Machine Intelligence**

MSc., Information Technology Management, Binary University, Puchong, Malaysia   Apr 2013 – Oct 2014
  - Specialization: **Information Technology**

BSc., Computer Science, University of Zambia, Lusaka, Zambia   Apr 2007 – Aug 2011
  - Specialization: **Computer Systems Engineering**

EXPERIENCE

Lecturer, Department of Computer Science, University of Zambia, Zambia   May 2015 – Feb 2024
  - **Bachelor of Computer Science courses:**

CSC 2011 - Computer Systems	2022/2023
CSC 2702 - Databases & Information System	2022/2023
CSC 3712 - Advanced Databases	2022/2023
CSC 4792 - Data Mining and Warehousing	2022/2023
CSC 4000 - Final Year Project Supervision	2022/2023

  - **Postgraduate Diploma in Computer Science courses:**

CSC 5711 - Databases & Information System	2022/2023

Instructor, Barbara Munyati CISCO Academy, University of Zambia, Zambia   Sep 2016 – Feb 2024
  - Instructor of Python Programming,
  - IT Essentials and Cisco Certified Network Associate (CCNA) Level 1 and Level 2 CISCO courses.

Teaching Assistant, Department of Computer Science, University of Zambia, Zambia   Feb 2012 – May 2015
  - Teaching assistant (Tutor) for Database Systems courses under the Staff Development Fellowship (SDF).

PERSONAL DEVELOPMENT

  - **Mathematics for Machine Learning Specialization** by Imperial College London on Coursera   Nov 2019
    Certificate of completion

  - **Deep Learning Specialization** by Deeplearningai on Coursera   Aug 2018
    Certificate of completion

  - **Machine Learning** by Stanford University on Coursera   Jul 2018
    Certificate of completion

PUBLICATIONS

- **Claytone Sikasote**, Kalinda Siaminwe, Stanly Mwape, Bangiwe Zulu, Mofya Phiri, Martin C. Phiri, David Zulu, Mayumbo Nyirenda, Antonios Anastasopoulos. Zambezi Voice: A Multilingual Speech Corpus for Zambian Languages, In *Proceedings of INTERSPEECH 2023 (pp. 3984-3988)*.

- **Claytone Sikasote**, Eunice Mukonde, Md Mahfuz Ibn Alam, Antonios Anastasopoulos. BIG-C: a Multimodal Multi-Purpose Dataset for Bemba, In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)(pp. 2062–2078)*.

- Odunayo Ogundepo,Tajuddeen R. Gwadabe, Clara E. Rivera, Jonathan H. Clark, Sebastian Ruder, David Ifeoluwa Adelani, Bonaventure F. P. Dossou, Abdou Aziz DIOP, **Claytone Sikasote**, Gilles Hacheme, Happy Buzaaba, Ignatius Ezeani, Rooweither Mabuya, Salomey Osei, Chris Emezue, Albert Njoroge Kahira, Shamsuddeen Hassan Muhammad, Akintunde Oladipo, Abraham Toluwase Owodunni, Atnafu Lambebo Tonja,Iyanuoluwa Shode, Akari Asai, Tunde Oluwaseyi Ajayi, Clemencia Siro, Steven Arthur, Mofetoluwa Adeyemi, Orevaoghene Ahia, Anuoluwapo Aremu, Oyinkansola Awosan, Chiamaka Chukwuneke, Bernard Opoku, Awokoya Ayodele, Verrah Otiende, Christine Mwase, Boyd Sinkala, Andre Niyongabo Rubungo, Daniel A. Ajisafe, Emeka Felix Onwuegbuzia, Habib Mbow, Emile Niyomutabazi, Eunice Mukonde, Falalu Ibrahim Lawan, Ibrahim Said Ahmad, Jesujoba O. Alabi, Martin Namukombo, Mbonu Chinedu, Mofya Phiri, Neo Putini, Ndumiso Mngoma, Priscilla A. Amuok, Ruqayya Nasir Iro, Sonia Adhiambo. AfriQA: Cross-lingual Open-Retrieval Question Answering for African Languages, In *Findings of the Association for Computational Linguistics: Empirical Methods in Natural Language Processing (EMNLP)(pp 14957–14972), 2023*

- **Claytone Sikasote**, Antonios Anastasopoulos. BembaSpeech: A Speech Recognition Corpus for the Bemba Language, In *Proceedings of the Thirteenth Language Resources and Evaluation Conference(LREC)(pp.7277–7283), 2022*.

- J. Kreutzer, I. Caswell, L. Wang, A. Wahab, D.V Esch, N. Ulzii-Orshikh, A. Tapo, N. Subramani, A. Sokolov, **C. Sikasote**, M. Setyawan, S. Sarin, S. Samb, B. Sagot, C. Rivera, A. Rios, I. Papadimitriou, S. Osei, P.O Suarez, I. Orife, K. Ogueji, A. Rubungo, T. Q. Nguyen, M. Müller, A. Müller, S.H. Muhammad, N. Muhammad, A. Mnyakeni, J. Mirzakhalov, T. Matangira, C. Leong, N. Lawson, S. Kudugunta, Y. Jernite, M. Jenny, O. Firat, B.F.P Dossou, S. Dlamini, N. Silva, S.Ballı, S. Biderman, A. Battisti, A. Baruwa, A. Bapna, P. Baljekar, I. A. Azime, A. Awokoya, D. Ataman, O. Ahia, S. Agrawal, M. Adeyemi. Quality at a glance: An audit of web-crawled multilingual datasets, *Journal of Transactions of the Association for Computational Linguistics (2022) 10: 50–72*.

- Simon H. Chiwamba, Jackson Phiri, Phillip O.Y. Nkunika, **Claytone Sikasote**, Monde M. Kabemba, Miyanda N. Moonga. Automated Fall Armyworm (Spodoptera frugiperda, J.E. Smith) Pheromone Trap Based on Machine Learning, *Journal of Computer Science 2019, 15 (12): 1759.1779*

- Leena Jaganathan, **Claytone Sikasote**, Mayumbo Nyirenda, Jackson Phiri. A Survey on Face Detection and Recognition Techniques for Application in Educational Institutions, In *Proceedings of the Zambia Association of Public Universities and Colleges (ZAPUC) International Conference, 2018*

SCHOLARSHIPS AND FELLOWSHIPS

- **Hasso-Plattner Institute (HPI) at University of Cape Town (UCT)**, Cape Town, South Africa   Nov 2022
- **African Master of Machine Intelligence Scholarship**, Kigali, Rwanda   Sep 2019
- **Special Research Fellowship**, University of Zambia, Lusaka, Zambia   Sep 2019
- **Binary University Scholarship**, Binary University, Malaysia   Apr 2013
- **Staff Development Fellowship**, University of Zambia, Lusaka, Zambia   Feb 2012
- **Zambia Higher Education Scholarship**, University of Zambia, Lusaka, Zambia   Apr 2007

GRANTS AND AWARDS

- **Recipient**, Lacuna Fund-Travel Grant, INTERSPEECH 2023   Aug 2023
- **Recipient**, ISCA-Travel Grant, INTERSPEECH 2023   Aug 2023
- **Recipient**, Diversity and Inclusion Virtual Award, ACL 2023   Jun 2023
- **Co-PI**, Multimodal Datasets for the Bemba Language Project   Apr 2021 - Aug 2022

TECHNICAL SKILLS

- **Programming Languages:** Python
- **Database technologies:** MS SQL, MySQL
- **Operating System platforms:** Windows, Linux
- **Machine Learning libraries:** Scikit-learn, Pandas, NumPy, SciPy
- **Data Visualization tools:** Matplotlib, tSNE
- **Deep Learning frameworks:** Pytorch, Tensorflow
- **Speech Recognition Tools:** DeepSpeech, FairSeq

OPEN-SOURCE DATASETS

- **Zambezi Voice** : [Lead researcher and creator] of a multilingual speech corpus containing speech recognition data resources for native African languages of Zambia. [ON-GOING]

  Link: *https://github.com/unza-speech-lab/zambezi-voice*

- **Multimodal Dataset for the Bemba** : [Lead researcher and Co-PI] of a large scale multimodal dataset for Bemba language containing multi-turn image-grounded dialogues amounting to over 180 hours of speech with corresponding Bemba transcriptions and translation in English. The project was funded by *Lacuna Fund*.

  Link: *https://github.com/csikasote/bigc*

- **BembaSpeech** : [Lead researcher and creator] an ASR corpus consisting of 24 hrs of read speech for the Bemba language spoken by over 30% of the population in Zambia.

  Link: *https://github.com/csikasote/BembaSpeech*

- **Multilingual ASR Corpus for African Languages** : [Contributor] to a multilingual speech corpus consisting of over 60 hours of read speech for 19 African languages. I contributed 2 hours of read speech for Bemba language.

  Link: *https://github.com/besacier/AMMIcourse/tree/master/STUDENTS-RETURN*

STUDENT SUPERVISION

- **John Mwiinga**, Undergraduate, University of Zambia   2022/2023
- **Nawiti Simwanza**, Undergraduate, University of Zambia   2022/2023
- **Felix Kunda**, Undergraduate, University of Zambia   2022/2023
- **Ngonidzashe Mbita Sikazwe**, Undergraduate, University of Zambia   2022/2023
- **Stanly Mwape**, Undergraduate, University of Zambia   2021/2022
- **Bangiwe Zulu**, Undergraduate, University of Zambia,   2021/2022
- **Kalinda Siaminwe**, Undergraduate, University of Zambia,   2021/2022

SERVICE

- **Committee Member, Department of Computer Science, University of Zambia**
  Projects and Exhibitions, Results Processing, Postgraduate Committee   2022/2023

- **Co-organizer, Dialectal and Low-resource Language Task, Bemba Language**
  International Conference on Spoken Language Translation (IWSLT 2024)   2023/2024

- **Programme Committee Member**
  Second Workshop on Computation and Written Language (CAWL 2024)   2023/2024
  ACL Special Interest Group on Writing Systems and Written Language (SIGWrit)   2023/2024

- **Conference Reviewer**
  Pan African Conference on Science, Computing & Telecommunications (PACT)   Sep 2023

- **Conference Volunteer**
  International Conference on Learning Representations (ICLR 2021)   Mar 2021
  International Conference on Artificial Intelligence and Statistics (AISTATS 2021)   Apr 2021